DOCUMENT RESUME

ED 441 855 TM 030 890

AUTHOR Wu, Yuh-Yin; Guei, I-Fen
TITLE Classroom Assessment Forms and Their Relations with Cognitive Components: An Example from Taiwan.
PUB DATE 2000-04-00
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Processes; Cognitive Tests; Correlation; *Elementary School Students; Foreign Countries; Intermediate Grades; Multiple Choice Tests; *Performance Based Assessment; Test Content; *Test Format
IDENTIFIERS Short Answer Tests; Taiwan

ABSTRACT

A study was conducted to investigate: (1) the relationships between the results from various forms of assessment and the patterns of correlation across content areas; (2) how cognitive components correlate with the test results from different classroom assessments; and (3) how content areas affected the relationships. Data were collected from a sixth-grade classroom of 40 students. Three assessment forms were administered crossing the two content areas of making neutral solutions and designing momentum experiments. The assessment forms were: (1) performance based assessment (PBA); (2) multiple choice (MC); and (3) short answer (SA). Findings were that MC and SA covaried more with one another than either did with PBA. Content area affected how PBA covaried with MC and SA. Deductive reasoning was the most obvious cognitive component that differentiated the designated content areas. This study demonstrated that PBA may not always be the best choice to measure students' cognitive capabilities. Paper and pencil tests may measure abstract relations as well as PBA does. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 855

Classroom Assessment Forms and Their Relations with Cognitive Components

—An Example from Taiwan

Wu, Yuh-Yin
Department of Educational Psychology and Counseling
National Taipei Teachers College, Taipei, Taiwan
Yuhyin@tea.ntptc.edu.tw

Guei, I-Fen
Le-Shing Elementary School, Taipei, Taiwan

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the Annual Meeting of American Educational Research Association, New Orleans, Louisiana, 2000.

Abstract

The purpose of the study was to investigate the following questions: 1. What were the relationships among the results from various forms of classroom assessment? 2. Were the patterns of correlation constant across content areas? 3. How did cognitive components correlate with the test results of different classroom assessment forms? 4. How did content areas affect the relationships? Data were collected from a 6th grade classroom of 40 students. Three assessment forms: performance based assessment (PBA), multiple choice (MC) and short answer (SA) were administered crossing two content areas: making neutral solution and designing momentum experiment. We found that MC and SA covaried more with one another than either did with PBA. Content area affected how PBA covaried with MC and SA. Deductive reasoning is the most obvious cognitive component that differentiated the designated content areas. This study demonstrated that PBA may not always be the best choice to measure students' cognitive capabilities. Paper and pencil tests may measure abstract relations as well as PBA does.

TM030890

BEST COPY AVAILABLE

Objectives

There are various forms of classroom assessment designed to measure students' abilities. How the results of various forms of assessment correlate with cognitive components should provide an idea about the interactive effects between the assessment approaches and the constructs being measured. The purpose of this study was then to investigate the following questions: 1. What were the relationships among test results from various forms of classroom assessment? 2. Were the patterns of correlation constant across content areas? 3. How did cognitive components correlate with the test results of different classroom assessment forms? 4. How did content areas affect the relationships between cognitive components and test results of different assessment formats?

Rationale

In Taiwan, objective paper-pencil test (PPT) is the main procedure for classroom assessment. Performance-based assessment (PBA) has been advocated since the educational reform in 1995. However, PPT is still widely used because it is less expensive comparatively. PPT mostly contains multiple-choice items, fill-in-the-blanks, and short answer questions. Classroom teachers usually try to stimulate students' performance on PPT by providing test items composed of basic recall plus analytic and synthesis non-routine problems. Yet, due to the push for PBA in educational reform trend, some teachers have modified their traditional forms of assessment by incorporating PBA along with PPT .

PBA has been emphasized in the United States since 1989 when Wiggins proposed authentic assessment (1989). The main characteristics of PBA are contextualized assessment procedures and higher order thinking skills. Baxter, Shavelson, Goldman, and Pine (1992) investigated correlations among PBA, CAT (cognitive ability test—paper-pencil form), and CTBS (comprehensive test of basic skills, science—paper-pencil form). They pointed out that the medium sized correlation between CTBS and PBA implied that the two assessment procedures *"tap[ped] different aspects of science knowledge and skills."* Furthermore, they found CAT and CTBS were more highly correlated than was either CAT or CTBS with PBA. Baxter et. al. (1992) concluded that this was due to the broad and abstract nature of CAT and CTBS versus the concrete nature of PBA.

For the aforementioned research findings, two assessment components—content area and assessment forms were simultaneously confounded in the correlation coefficients. PBA used a

different assessment form and covered different content areas from that of CAT and CTBS. Therefore, the low correlation between CAT and PBA could be due to either the different assessment forms or different content areas. Thus this current study was designed to investigate further how different assessment forms across different content areas would correlate. This design should help classroom teachers and researchers in understanding how similar or different the various classroom assessment forms are when controlling the content areas. Furthermore, we attempted to investigate the relationships of test results with cognitive components. The intention was to identify what was being measured by the different assessment forms from various content areas.

Method

Context The study was conducted at a public school in the Taipei metropolitan area in Taiwan. A 6th grade class was selected which consisted of 40 students. The content areas of the test were 'making neutral solution (Sol)' and 'designing momentum experiment (Mom)' within the subject of physical sciences. For each content area, students were assessed by performance-based assessment (PBA), multiple choice items (MC), and short answer questions (SA). PBA took place two days before MC and SA. MC and SA were administered at the same hour session with MC being given first. All the tests were administered during fall semester, 1996.

PBA (Performance-Based Assessment) For the task of SolPBA (performance-based assessment to make neutral solution), students were given 4 kinds of liquid of different pH values and the necessary laboratory equipment. They were required to make a neutral solution and to provide the ratio of each liquid used. The scoring rubrics contained 20 rating items covering method, procedures, results, and interpretation. For the task of MomPBA (PBA to measure momentum), students were asked to design an experiment (as depicted in Figure 1) by creating hypotheses, and manipulating variables while holding the other variables constant. Students were required to draw conclusions about how variables affected momentum. The scoring rubrics contained 16 rating items covering design, procedures, results, and interpretation. The second author was the students' physical sciences teacher and served as one of the two raters. The second rater was also a physical sciences teacher employed at the same school.

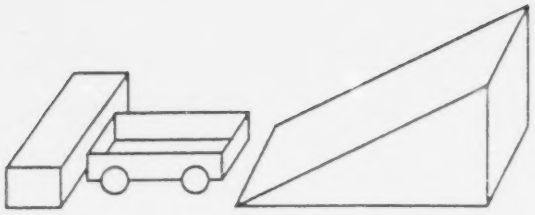

Figure 1. Momentum Experimental Design Outlook

PPT (Objective Paper-Pencil Test) Paper-pencil tests from both Sol and Mom were created and were composed of two sessions: multiple choice items (MC) and short answer questions (SA). The items were content-validated using Bloom's cognitive taxonomy table (1956). Sol paper-and-pencil test consisted of 25 MC items and 17 SA items while Mom paper-and-pencil test consisted of 25 MC items and 9 SA items.

Ross Test of Higher Cognitive Processes The Ross Test was designed to assess higher level thinking skills. It contained 8 cognitive components (as shown in Table 1) based on Bloom's (1956) definition of cognitive hierarchy.

Table 1. The Eight Cognitive Components of Ross Test of Higher Cognitive Processes

Cognitive Components	Corresponding Acronym
Analysis of attributes	AA
Analysis of relevant and irrelevant information	AI
Analogies	AN
Abstract relations	AR
Deductive reasoning	DR
Missing premises	MP
Questionning strategies	QS
Sequential synthesis	SS

The Ross Test was given at the end of the spring semester, 1997. It was chosen as the measure of cognitive components primarily because it was one of the few available cognitive tests with a Chinese version. The Ross Test was translated into Chinese and standardized by Lin, Jen, Guo, and Fang (1991). The stability coefficients ranged from .42 to .76, and internal consistency ranged from .43 to .75. The criterion validity coefficients with school achievements ranged from .10 to .76. According to Ross and Ross (1976), the split-half reliability coefficient of the test was .93. Evidence of the Ross Test validity was low correlation with the Lorge-Thorndike Intelligence Test and significant correlation with

chronological age.

Results

Crossing 2 content areas and 3 assessment forms, there were 6 assessment procedures forming a multi-trait multi-method correlation matrix as shown in Table 2.

Table 2. MTMM matrix of six assessment procedures (N=40)

	Sol (Making Neutral Solution)			Mom (Measuring Momentum)		
	SolMC	SolSA	SolPBA	MomMC	MomSA	MomPBA
	25 items	17 items	2 raters	25 items	9 items	2 raters
SolMC	.703**1					
SolSA	.675**	.804**1				
SolPBA	.481**	.462**	.97**2			
MomMC	.462**	.505**	.309	.742**1		
MomSA	.414**	.656**	.294	.765**	.84**1	
MomPBA	.651**	.689**	.399	.283	.471**	.97**2

1: internal consistency

2: inter-rater correlation

** Correlation is significant at the .01 level (2-tailed).

1. In the mono-trait hetero-method convergent validity areas as noted by two triangles, MC correlated higher with SA than with PBA (.675>.481 and .765>.283). In the content area of Sol, SolPBA was correlated about the same with SolMC and SolSA (.481 and .462). In the content area of Mom, MomPBA correlated lower with MomMC than with MomSA (.283). Of all the coefficients in the matrix the lowest coefficient was found between MomMC and MomPBA (.283<.471). This is a convergent validity coefficient which ideally would be high. The results indicated that MC and SA were more similar and their size of correlations with PBA were dependent upon different content areas (Sol or Mom) . MC and SA shared more variation with PBA in Sol which contained more chemical declarative knowledge than they did with PBA in Mom which contained more experimental design procedural knowledge.

2. In the hetero-trait discriminant validity coefficie t area as noted by a rectangle in Table 2, coefficients were theorectically assumed to be relatively lower than the convergent validity coefficients as indicated by the triangles. Hetero-trait mono-method coefficients (diagonal coefficients) were supposed to be higher than their corresponding row and column off diagonal

coefficients (hetero-trait hetero-method coefficients). However, the facts in the study were as follows:

(1) The three diagonal correlation coefficients (.462, .656, and .399) were not at all the highest same method coefficients in the rectangle.

(2) In the row of MomPBA, the correlation between MomPBA and SolPBA should be the highest while it turned out to be the lowest. (.399< .651 or .689)

(3) MomPBA correlated highly with SolMC and SolSA (.651 and .689) while MomSA correlated highly with SolSA (.656). These discriminant coefficients were higher than six of the seven corresponding convergent coefficients (as indicated in the triangles).

The generalizability coefficient of PBA across 2 content areas and 2 raters was 0.57, similar to the size reported by Ruiz-Primo, Gaxter, and Shavelson (1993). The source of variation from the interaction between students and content areas was 56.37%, from students' within subject variation was 38.78%, and from the unexplained error source was 4.85%. These proportions indicated the influence of content areas in rating students' performance. This finding supported the fact that PBA is contexualized and content dependent.

3. To further analyze the characteristics these classroom assessment forms measure, eight cognitive components devised by Ross and Ross (1976) were adopted to correlate with the classroom test results as shown in Table 3.

Table 3. The correlation between the eight cognitive components and the test results from six assessment forms (N=40).

	AA	AI	AN	AR	DR	MP	QS	SS	Total
SolMC	.228	.458**	.590**	.445**	.698**	.428**	.421**	.315	.649**
SolSA	.140	.439**	.536**	.626**	.512**	.568**	.426**	.366	.645**
SolPBA	.336	.463**	.470**	.316	.541**	.373	.409**	.377	.610**
MomMC	.126	.364	.249	.296	.380	.360	.311	.469**	.450**
MomSA	.267	.339	.316	.331	.319	.603**	.354	.308	.504**
MomPBA	.011	.420**	.529**	.524**	.388	.556**	.497**	.239	.569**

Of the 8 cognitive components, deductive reasoning (DR) differentiated test results of Sol from Mom (.698, .512, and .541 vs. .380, .319, and .388). As mentioned MomPBA shared more common variance with SolMC and SolSA than with other assessment forms. These three assessment forms also simultaneously and exclusively correlated higher with AR than

with the other cognitive components. Test results from assessment forms in Sol content area tended to correlate higher with various cognitive components, particularly with AI , AN, DR, and QS. In terms of Mom, test results from MomMC and MomSA tended to have lower correlation coefficients with the cognitive components.

Discussion

This study demonstrated that within individual content area test results of multiple choice (MC) and short answer (SA) covaried more with each other than with performance-based assessment (PBA). It was assumed that under the same content area, MC and SA tapped similar cognitive components while PBA emphasized different cognitive dimensions. When crossing two different content areas, SolPBA and MomPBA did not show stronger association with each other. Generalizability analysis indicated that same PBA tasks could also measure very different traits due to content area. In this study, Sol required students to know very clearly the definition and properties of acid and base liquids before they proceeded. Chemical knowledge was emphasized. Mom required students to conduct experiments, particularly controlling and manipulating variables. Procedure skills were more emphasized in the task.

It was found that deductive reasoning is the most obvious cognitive component that distinguished the content areas. Deductive reasoning also made Sol content, compared to Mom content, easier to measure by objective paper-and-pencil test. That was why we saw that SolMC, SolSA, and SolPBA shared more common cognitive components. This was also confirmed by the minimal relationship between MomMC and MomPBA although they were from the same content area. The characteristics of Mom caused difficulties in composing MC and SA assessment forms. It resulted in low coefficients of these forms with various cognitive components.

Interestingly, the correlations of test results showed that MomPBA was more similar with SolMC and SolSA from different content area than with MomMC and MomSA under the same content area. The cognitive component of abstract relations (AR) correlated with the three assessment forms, MomPBA, SolMC, and SolSA more significantly than with the other three forms, SolPBA, MomMC, and MomSA. It was found that AR was emphasized in MomPBA task when students tried to manipulate one variable while controlling the other variables or when they manipulated different variables to detect the influential factors of momentum. The process of deciding cause and effect relationships was similarly emphasized in SolSA or SolMC test items when students were asked items such as *"Tom felt acid in his mouth when*

he hiccuped. Then which food should Tom take less at meals?" and *" Which drink can turn lemonade neutral?"* The common empahsis of AR in MomPBA, SolCM, and SolSA could the factor that associated these three assessment forms together.

Educational Implications

The similarities and differences of test results from multiple choice items (MC), short answers (SA), and performance based assessment (PBA) reminded us of Linn and Gronlund' s concern that *"performance assessment should be on measuring complex achievement that can not be measured well by objective tests."* (1995, p. 261). This study demonstrated for teachers and researchers that PBA is not always the best choice for measuring students' cognitive capabilities. PPT may share a large amount of variance with PBA even across content areas. The cognitive components, such as abstract relations, could be measured well by SA. Our results indicated that ability in neutral solution content can be measured well by MC, SA, and PBA, while momentum experiment is not suitable for MC and SA. Therefore, it is essential that assessment forms be selected based upon content area and the traits of cognitive components so that the validity and its cost-efficiency can be maximized.

References

Baxter, G. P., Shavelson, R. J., Goldman, S. R. and Pine, J. (1992). Evaluation of procedure-based scoring for hands-on science assessment. *Journal of Educational Measurement, 29*(1), 1-17.

Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objective, hand I: The cognitive domain.* New York: McKay.

Lin, L. - F., Jen, Y.-M., Guo, S.-Y. and Fang, Y.-M. (1991). *Chinese Ross test of higher cognitive processes manual.* Taipei: Teachers Association of Taipei.

Linn, J. E. and Gronlund, M. A. (1995). *Measurement and assessment in teaching* (7th Ed.). New Jersey: Prentice-Hall.

Ross, J. and Ross, C. (1976). *Ross test of higher cognitive processes manual.* Novato, CA: Academic Therapy Publication.

Ruiz-Primo, M. A., Baxter, G. P. and Shavelson, R. J. (1993). On the stability of performance assessments. *Journal of Educational Measurement, 30*(1), 41-53.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 20*, 703-713.